Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-World Concordance of the European Association of Urology Guidelines Bot With Multidisciplinary Decision-Making in Prostate Cancer: A Prospective Study

Alfonso Romero Crespo

alfonromero@usal.es

Hospital Regional Universitario de Málaga

Javier Fernandez Siles

Hospital Regional Universitario de Málaga

Rodrigo España Navarro

Hospital Regional Universitario de Málaga

Short Report

Keywords: Artificial intelligence, Clinical decision support, Clinical guidelines, Multidisciplinary tumor board, Prostate cancer

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11446690/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Guideline-based artificial intelligence tools may support oncological decision-making, but their performance in real-world multidisciplinary care remains uncertain. We evaluated the concordance between recommendations generated by the European Association of Urology (EAU) Guidelines Bot and

Findings:

We prospectively evaluated 156 patients with prostate cancer discussed at a multidisciplinary tumor board between June 2025 and April 2026 using a standardized clinical prompt. Recommendations were classified as complete concordance (level 1), correct but incomplete (level 2), partially correct with erroneous recommendations (level 3), or completely incorrect (level 4). Complete concordance was observed in 74.3% of cases, while 89.7% were considered clinically acceptable (levels 1–2). Overall, 39 cases showed some degree of discrepancy (levels 2–4). The main sources of discrepancy were lack of awareness of local clinical trials or protocols (20.5%), inadequate adaptation to frailty or severe comorbidity (15.4%), omission of genetic testing (10.3%), and omission of required diagnostic or restaging steps (10.3%).

Conclusions

The EAU Guidelines Bot showed high concordance with multidisciplinary decision-making, but limitations emerged when recommendations required patient-specific or locally contextualized information. These findings support its use as a decision-support tool rather than a substitute for multidisciplinary clinical judgment.

Introduction

Multidisciplinary tumor boards (MTBs) are central to complex oncological decision-making, integrating clinical, pathological, imaging, and patient-specific factors. Artificial intelligence–based decision-support tools may facilitate guideline implementation, but their performance in real-world clinical decision-making requires validation. Recent prospective evidence in urological oncology has shown substantial, although imperfect, agreement between large language model recommendations and MTB decisions. The EAU Guidelines Bot represents a particularly relevant tool for evaluation because it is specifically designed around EAU guideline content. We therefore prospectively evaluated its concordance with MTB decisions in patients with prostate cancer and explored the clinical scenarios underlying discordant recommendations.

Methods

We prospectively evaluated 156 patients with prostate cancer discussed at a multidisciplinary tumor board (MTB) at a tertiary referral center between June 2025 and April 2026. Clinical information available for each case was entered into the EAU Guidelines Bot using an identical standardized prompt incorporating age, comorbidities and frailty, relevant medical and oncological history, prostate-specific antigen level and doubling time when applicable, histopathology, TNM stage, metastatic burden, and germline/somatic testing (Supplementary Table 1).

The Bot was instructed to provide a primary guideline-concordant recommendation, valid alternatives, and a brief patient-specific rationale. MTB consensus served as the clinical reference standard. MTB members were unaware of the study comparison and had no access to Bot outputs when reaching their decisions. Bot recommendations were independently assessed by two reviewers using a predefined four-level scale: level 1, correct and complete; level 2, correct but incomplete; level 3, partially correct with erroneous recommendations; and level 4, completely incorrect. Disagreements were adjudicated by a third reviewer. Level 1 represented complete concordance, while levels 1–2 were considered clinically acceptable. Cases with incomplete or discordant recommendations (levels 2–4) underwent a structured post hoc analysis to characterize the underlying causes. Descriptive analyses evaluated concordance and patterns of discordance overall and according to disease stage and proposed therapeutic modality. Agreement was quantified using Cohen's κ.

Results

Among 156 patients, complete concordance (level 1) between the EAU Guidelines Bot and MTB assessment was observed in 74.3%, while 89.7% of recommendations were considered clinically acceptable (levels 1–2). Accordingly, 39 cases showed some degree of incompleteness or discordance (levels 2–4). Agreement was substantial when complete concordance was required ($\kappa = 0.60$) and increased when levels 1–2 were considered clinically acceptable ($\kappa = 0.85$).

Structured post hoc analysis of the 39 level 2–4 cases identified lack of awareness of local clinical trials or institutional protocols as the most frequent source of discrepancy (8/39, 20.5%), followed by inadequate adaptation to frailty or severe comorbidity (6/39, 15.4%), omission of genetic or molecular testing (4/39, 10.3%), and omission of required diagnostic or restaging procedures (4/39, 10.3%). Discordance patterns differed according to disease stage. Among the eight discordant metastatic cases, omission of genetic or molecular testing accounted for 50.0% (4/8), whereas locally dependent recommendations and additional diagnostic or restaging requirements were confined predominantly to localized or locoregional disease. Discordances were also observed across different therapeutic scenarios, indicating that clinically acceptable overall concordance did not eliminate context-dependent limitations (Table 1).

Table 1
Causes of discordance according to prostate cancer stage

Cause of discordance	Global, n (%)	Localized/locoregional, n (%)	Metastatic, n (%)
No. of patients	156 (100%)	124 (79.5%)	32 (20.5%)
Discordant cases	39 (100%)	31 (79.5%)	8 (20.5%)
Unawareness of local clinical trials or protocols	8 (20.5%)	8 (25.8%)	0 (0.0%)
Failure to account for frailty and severe comorbidities	6 (15.4%)	4 (12.9%)	2 (25.0%)
Omission of genetic or molecular testing	4 (10.3%)	0 (0.0%)	4 (50.0%)
Omission of required diagnostic procedures (staging failure)	4 (10.3%)	4 (12.9%)	0 (0.0%)
Failure to account for toxicity risk, contraindications, or previous treatments	4 (10.3%)	4 (12.9%)	0 (0.0%)
Failure to consider PSA kinetics and/or potential overtreatment	3 (7.7%)	3 (9.7%)	0 (0.0%)
Failure to integrate extra-urological findings (synchronous malignancies)	3 (7.7%)	3 (9.7%)	0 (0.0%)
Inability to directly assess medical imaging	2 (5.1%)	1 (3.2%)	1 (12.5%)
Nonspecific recommendation or multiple options without prioritization	2 (5.1%)	2 (6.5%)	0 (0.0%)
Omission of supportive therapies or specific histological follow-up	2 (5.1%)	1 (3.2%)	1 (12.5%)
Other methodological errors or nonspecific omissions	1 (2.6%)	1 (3.2%)	0 (0.0%)

Percentages for causes of discordance were calculated using discordant cases as the denominator: n = 39 overall, n = 31 for localized/locoregional disease, and n = 8 for metastatic disease. One primary cause of discordance was assigned to each case. PSA: prostate-specific antigen.

Discussion

This prospective study demonstrated complete concordance between the EAU Guidelines Bot and multidisciplinary tumor board (MTB) decisions in 74.3% of prostate cancer cases, increasing to 89.7% when clinically acceptable responses were considered. These findings align with Nardone et al. [1], who highlighted the potential of AI to optimize MTB workflows while noting variability in performance across

specialties. Although direct comparisons are limited by differences in case mix, AI systems, and concordance definitions, de la Torre-Torres et al. [2] reported 56.1% complete agreement using ChatGPT-4o in urological tumors (κ = 0.61). The higher complete concordance observed in our cohort may reflect the use of a domain-specific tool grounded in EAU guideline content.

However, concordance alone does not establish the ability to replace multidisciplinary decision-making. Our structured analysis of the 39 level 2–4 cases provides particular insight into where this approach fails. The most frequent source of discordance involved local clinical trials or institutional protocols (20.5%). Guideline-based and general-purpose AI systems cannot be assumed to incorporate center-specific pathways, emphasizing the need for local validation before clinical implementation [1]. A second limitation is the inability to directly assess medical imaging—a cornerstone of oncological decision-making—requiring reliance on text-based descriptions that may not capture subtle radiological findings.

More broadly, human specialists integrate information beyond guideline recommendations, including frailty, severe comorbidity, previous treatments, PSA kinetics, imaging findings, and patient preferences. This was reflected in our cohort, where inadequate adaptation to frailty or severe comorbidity accounted for 15.4% of discordances. The omission of genetic or molecular testing in 4/8 discordant metastatic cases further illustrates how an otherwise reasonable recommendation may remain clinically incomplete.

These limitations raise an important distinction between AI-assisted and autonomous clinical decision-making. Given the potential consequences of incomplete or contextually inappropriate recommendations, AI outputs should remain subject to expert oversight rather than being considered independently actionable therapeutic decisions [3]. As Albisua and Pacheco [4] emphasized, current ethical frameworks support AI as a complement to, rather than a replacement for, clinical judgment, with human oversight remaining essential. Therefore, the EAU Guidelines Bot should currently be regarded as a decision-support tool for standardizing guideline-based care rather than a substitute for multidisciplinary clinical judgment.

Conclusions

In conclusion, the EAU Guidelines Bot showed high concordance with MTB decision-making in prostate cancer, particularly for guideline-driven scenarios. However, discrepancies emerged when recommendations required integration of patient-specific factors, local resources, or additional diagnostic information. These findings support its potential as a guideline-based decision-support tool while reinforcing the continued need for multidisciplinary clinical judgment.

Abbreviations

- **AI:** Artificial Intelligence

- **CHAARTED:** Chemohormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease in Prostate Cancer
- **EAU:** European Association of Urology
- **ECOG:** Eastern Cooperative Oncology Group
- **ISUP:** International Society of Urological Pathology
- **MTB:** Multidisciplinary Tumor Board
- **PSA:** Prostate-Specific Antigen
- **PSADT:** Prostate-Specific Antigen Doubling Time
- **TNM:** Tumor, Node, Metastasis

Declarations

Our study was an observational investigation involving anonymized clinical data from patients with prostate cancer. No experimental interventions were performed, and clinical management was not influenced by the study. All patients had signed informed consent authorizing the processing and use of their clinical data. The study was discussed with the Ethics Committee of Hospital Regional Universitario de Málaga, which indicated that formal ethics approval was not required given the observational nature of the study and the use of anonymized data.

Ethics approval and consent to participate:

This study was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all patients. The protocol was conducted using observational registry data within the multidisciplinary tumor board framework.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

The authors declare that no financial support was received from any organization for the submitted work.

Author Contribution

A.R.C., J.F.S., and R.E.N. conceived and designed the study. A.R.C. and J.F.S. acquired the data. A.R.C., J.F.S., and R.E.N. analyzed and interpreted the data. A.R.C. drafted the main manuscript text, and J.F.S.

and R.E.N. critically revised it for important intellectual content. All authors reviewed and approved the final manuscript.

Acknowledgements:

Not applicable.

Authors' information:

Not applicable.

Availability of data and materials:

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Nardone V, Marmorino F, Germani MM, Cichowska-Cwalinska N, Menditti VS, Gallo P, et al (2024) The Role of Artificial Intelligence on Tumor Boards: Perspectives from Surgeons, Medical Oncologists and Radiation Oncologists. Curr Oncol 31(9):4984-5007
2. de la Torre-Torres JDI, Carrasco YY, Segura MTM, Sánchez EC, Muñoz AZ, Morales-Domínguez J, et al (2026) Concordance Between ChatGPT-4o and a Multidisciplinary Urologic Oncology Tumor Board: A Real-World Clinical Decision-Making Study. J Clin Med 15(6):2130
3. Ammo T, Guillaume VGJ, Hofmann UK, Ulmer NM, Buenting N, Laenger F, et al (2024) Evaluating ChatGPT-4o as a Decision Support Tool in Multidisciplinary Sarcoma Tumor Boards: Heterogeneous Performance across Various Specialties. Front Oncol 14:1526288
4. Albisua J, Pacheco P (2025) Ethical Considerations About Artificial Intelligence (AI) Use in Medical Practice. Open Respir Arch 7:100383

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.docx